Transcriptionally Silenced Plant Genes

The present invention relates to the field of gene expression in plants and in particular concerns gene silencing, a phenomenon frequently observed after integration of transgenes into plant genomes. Comparison of transcriptional gene expression between an Arabidopsis line carrying a silent transgene present in multiple copies and its mutant derivative *mom1* impaired in silencing of the transgene revealed two cDNA clones which are expressed in the mutant plants, but not in the parental and not in wild type plants. Both clones are derived from the same family of transcripts which we refer to as TSI (Transcriptionally Silent Information). The disclosed genomic templates encoding TSI are repetitive elements with mainly pericentromeric location and conserved organization among various ecotypes. They are also referred to as TSI. Transcriptional silencing of the genomic TSI templates is specifically released in the mutant. Silencing of said templates is further released in other genotypes known to affect transcriptional gene silencing. Thus, transcription of TSI can be used as a marker to identify a defective silencing pathway in a plant.

Correct balance between activation and silencing of its genetic information is essential for any living cell. A tight control of gene expression is necessary for adaptation to environmental factors, regulation of physiological requirements, and development of differentiated, specialized cell types within a multi-cellular organism. For example differentiation processes involve mitotically heritable changes of gene expression, wherein the acquired states of gene activity gain a certain stability. This stability can be achieved by the strict control of gene activators, by regulation of transcript stability, or by regulating the transcriptional availability of genetic information itself as by stable silencing of selected genetic loci. Silencing has been frequently observed in connection with repression of transgene expression in various experimental systems.

In plants, silencing of transgenic loci limits the reliability of transgenic approaches to improve quality traits. It has been noticed that complex inserts containing rearranged multiple copies of a transgene are particularly prone for gene silencing. Two different mechanisms leading to loss of transgene expression are observed. The first prevents transcription (transcriptional gene silencing or TGS), and the second targets selected transcripts for rapid degradation (posttranscriptional gene silencing or PTGS). Triggers of both processes seem to be similar, since the onset of both types of silencing correlates with redundancy of genetic information, i.e. DNA repeats in case of TGS and RNA

overproduction for PTGS. TGS is meiotically heritable and correlates with DNA template modification manifested by hypermethylation of promoters of silenced genes or with local changes of chromatin structure. In contrast, PTGS is not meiotically transmitted and needs to be reestablished in each sexual generation. PTGS does not require modification of a DNA template, however, increased levels of DNA methylation within the protein-coding region of silenced genes have been observed.
The majority of silencing studies in plant systems deal with silencing of transgenes. There are only a few examples of gene silencing without involvement of transgenic loci. The criteria for TGS susceptibility of genetic information is very poorly understood, and the natural targets of transcriptional silencing in a normal, wild type plant are yet to be discovered. It has been postulated that TGS is a defense system against invasive DNA such as transposable elements but experimental evidence for this hypothesis is lacking.

Within the context of the present invention reference to a gene is to be understood as reference to a DNA coding sequence associated with regulatory sequences, which allow transcription of the coding sequence into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Examples of regulatory sequences are promoter sequences, 5' and 3' untranslated sequences, introns, and termination sequences.
A promoter is understood to be a DNA sequence initiating transcription of an associated DNA sequence, and may also include elements that act as regulators of gene expression such as activators, enhancers, or repressors.
Expression of a gene refers to its transcription into RNA or its transcription and subsequent translation into protein within a living cell.
Any part or piece of a specific nucleotide or amino acid sequence is referred to as a component sequence.

It is the aim of the present invention to provide nucleic acid molecules encoding genetic information which is not expressed, i.e. silenced, in wild type plants but whose expression is turned on in plants which are defective in transcriptional gene silencing. Said molecules can be defined by the formula R_A-R_B-R_C, wherein

- R_A, R_B and R_C indicate component sequences consisting of nucleotide residues independently selected from the group of G, A, T and C or G, A, U and C, wherein
  G is Guanosinmonophosphate,
  A is Adenosinemonophosphate,

T is Thymidinmonophosphate,
U is Uridinmonophosphate and
C is Cytidinmonophosphate;

- R_A and R_C consist independently of 0 to 6000 nucleotide residues;
- R_B consists of at least 50 nucleotide residues; and
- the component sequence R_B is at least 80% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

In a preferred embodiment of the present invention R_B consists of at least 100 nucleotide residues and is at least 85% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.
In another preferred embodiment R_B consists of at least 200 nucleotide residues and is at least 90% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.
Specific examples of R_B are the sequences given in SEQ ID NO: 7, SEQ ID NO: 9 and SEQ ID NO: 27.
Additionally, R_A or R_C may comprise one or more component sequences with a length of at least 50 nucleotide residues and at least 90% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

The nucleic acid molecules according to the present invention exist either in the form of DNA or as RNA. Preferred embodiments are genomic DNA, cDNA, plasmid DNA or RNA transcribed therefrom.

Nucleotides 437-2383 of SEQ ID NO: 1 encode a putative open reading frame of 648 amino acids (SEQ ID NO: 10) which in SEQ ID NO: 1 is interrupted by a stop codon spanning nucleotides 1631-1633. Nucleic acids encoding a protein comprising a component sequence of at least 200 amino acids length being at least 85% identical to an aligned component sequence of SEQ ID NO: 10 are a further preferred embodiment of the present invention.

The nucleic acids according to the present invention represent an endogenous target of the transcriptional silencing system. Example 1 describes the cloning of specific embodiments of the present invention from Arabidopsis. The preferred size of transcribed RNA is between 1000 and 6000 nucleotides, particularly transcripts of about 1250, 2500, 4700 and 5000 nucleotides, which can be polyadenylated or not. The transcriptionally silent information present in the genome of wild type plants is found to be only expressed in a range of mutants affected in the maintenance of transcriptional silencing. Importantly, not only strains affected in transcriptional silencing through alterations of genome-wide DNA methylation, but also silencing mutants with unchanged methylation levels which do not show striking phenotypic alterations activate TSI, indicating that the release of silencing from endogenous templates does not require loss of methylation.
Initially two independent clones representing RNA which is specifically expressed in silencing mutants have been cloned. Anticipating that in wild type plants there are probably many more DNA templates suppressed by the silencing system, it is remarkable that parts of the two cDNAs cloned are closely related to each other and it is most likely that they are parts of the same transcript. The three main TSI transcripts of 5000, 2500 and 1250 nucleotides all contain a middle element isolated as TSI-A (SEQ ID NO: 5). The 5000 nt and the 2500 nt transcripts additionally enclose the second isolated element TSI-B (SEQ ID NO: 6), which is like TSI-A without protein coding capacity. The 5000 nucleotide long transcript further comprises a 5' extension (SEQ ID NO: 1 which is similar to SEQ ID NO: 2) encoding a putative open reading frame of 648 amino acids (SEQ ID NO: 10). The two 3' extension clones of TSI-A (SEQ ID NO: 3 and SEQ ID NO: 4) contain a region which can be aligned with nucleotides 1-569 of SEQ ID NO: 6 (nucleotides 808-1397 of SEQ ID NO: 3 and nucleotides 819-1365 of SEQ ID NO: 4) closely related to TSI-B (77 % identity).
Both the 5000 and the 2500 nucleotide transcripts are polyadenylated, while the most abundant transcript of 1250 nucleotides is absent from the polyA fraction of *mom1* RNA and might be retained in the nucleus.

All RNA species originate from unidirectional transcription, but it is not clear if they represent separate transcriptional units regulated by different promoters or if they are processing products of the same long transcript. A refined analysis of the TSI expression pattern is complicated by the multiplicity of potential chromosomal templates and their location mainly in the pericentromeric areas.

The novel TSI sequences do not reveal any putative function by sequence similarity to protein- or RNA-coding sequences. The only extensive similarity found was to the 3' half of the putative, degenerated retrotransposon *Athila* (Pélissier et al. 1995). The other part of *Athila* directly adjacent to the TSI template region was not reactivated in the silencing mutant. This suggests that epigenetic transcriptional silencing in *Arabidopsis* is not directed towards retrotransposons in general, although its targets may have originated from transposition events. This is further supported by the lack of transcriptional reactivation of other *Arabidopsis* retroelements, e.g. the *Ta* superfamily (Konieczny et al. 1991). Therefore, only specific pericentromeric repeats seem to be under epigenetic control, in the same way that only a subset of transgenic loci is susceptible to silencing. The existence of remnants of transposons is probably due to their chromosomal location rather than to sequence specificity, since degenerated retroelements have repeatedly been found in centromeric locations in fungi and plants.

One of the features proposed as a prerequisite for centromere function is late replication of the heterochromatic centromeres and pericentromeric areas in *Schizosaccharomyces pombe* and higher eukaryotes. If this was also true for *Arabidopsis* centromeres, undue loosening of suppressive chromatin leading to TSI expression could cause disturbances in mitosis, which would result in severe phenotypes. However, the *mom* mutant plants exhibit no abnormalities suggesting mitotic disorders. Therefore, transcriptional reactivation of some usually silent pericentromeric repeats, such as described here, does not impair their putative function. Alternatively, their silencing may be important under a specific, still undefined condition or on a longer time scale.

Finally, TSI expression is observed in cells growing for a long time in suspension culture. No release of TSI silencing is observed in any tissue of *Arabidopsis* wild type plants, including freshly initiated callus cultures. This suggests that an escape from the silencing control is not correlated primarily with dedifferentiation but could be the result of prolonged selection for fast growing dedifferentiated cells. Such a loss of silencing control could also underlie the accumulation of somaclonal variation during prolonged culture and resembles the situation in cells of actively proliferating carcinomas.

Nucleic acids according to the present invention are particularly useful in selecting plants which compared to wild type Arabidopsis plants of all available Arabidopsis ecotypes are

impaired in transcriptional gene silencing. A method allowing to select such plants comprises

a) separately preparing RNA of a series of plants;
b) probing said RNA preparations with a nucleic acid according to the present invention; and
c) identifying a plant whose RNA hybridizes with said nucleic acid.

In a preferred embodiment the probing step is performed after size fractionation of the RNA preparation by gel electrophoresis. For detection the probe is either radioactively labeled or labeled by other chemical modifications.

In another preferred embodiment of said method the step of probing consists of hybridizing the RNA with an oligonucleotide primer, extending said primer by reverse transcription and subsequent PCR amplification of the DNA generated using oligonucleotide primers specific for SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 27. Plants which allow for the amplification of DNA fragments flanked by the oligonucleotide primers are identified as plants whose RNA hybridizes to the nucleic acid according to the invention.

Having available nucleotide sequence information of a genomic region, which is not expressed, i.e. transcriptionally silenced, in a wild type plant, allows to produce DNA representing at least part of a gene necessary to maintain silencing of this genomic region. Preferably the complete gene is produced. A corresponding method of production comprises

(a) mutagenizing wild type cells or plants by randomly inserting into their genomes a DNA tag with known sequence;
(b) identifying mutants of said cells or plants which express RNA that is not expressed in wild type cells or plants;
(c) cloning genomic DNA surrounding or close to the insertion site of the DNA tag;
(d) screening a genomic library of wild type cells or plants with the piece of genomic DNA obtained in process step (c) or a part thereof;
(e) identifying clones comprising at least part of the gene affected by the insertion of the DNA tag; and
(f) further processing the clones obtained in step (e) using recombinant DNA techniques.

In plant cells and plants mutagenesis is preferably achieved performing T-DNA insertion mutagenesis (Dilkes and Feldmann, 1998), or transposon tagging using the En/I (Pereira and Aarts, 1998) or the Ac/Ds system (Long and Coupland, 1998) as described in Arabidopsis protocols edited by Martinez-Zapater and Salinas, 1998. Other known physical or chemical methods of mutagenesis such as fast neutron irradiation or EMS mutagenesis (Feldmann et al., 1994) might require adaptation of the above method, but can be used for the production of equivalent DNA involved in the maintenance of silencing as well.
A convenient way to identify RNA that is expressed in mutant cells or plants but not in wild type cells or plants is reverse transcription of said RNA and subsequent PCR amplification of the generated DNA using oligonucleotide primers specific for said DNA (RT-PCR). This allows to pool the RNA of upto 1000 mutants which increases the speed of the identification step considerably.

The methods described above can be further elaborated and developed into a kit for the identification of plants impaired in transcriptional gene silencing. Such a kit necessarily comprises

a) a nucleic acid according to the present invention conveniently labeled to be used as a hybridization probe or
b) an oligonucleotide primer for reverse transcription of RNA and an oligonucleotide primer specific for a nucleic acid according to the present invention.

The oligonucleotide primer for reverse transcription can be a poly T primer or an oligonucleotide primer specific for a nucleic acid according to the present invention. The primers specific for nucleic acids according to the present invention are designed to allow PCR amplification of DNA templates characterized by the nucleotide sequences disclosed in the present invention.

EXAMPLES

Example 1: **Differential mRNA screening and cloning of *Arabidopsis* TSI sequences**

Total RNA of the mutant line *mom1* (Amedeo et al, 2000) and its parental line A is isolated according to Goodall et al. (1990) using 2 g fresh weight of two-week-old seedlings. Polyadenylated RNA is obtained using Dynabeads Oligo $(dT)_{25}$ (Dynal). 2 µg of poly(A) RNA is used for suppression subtractive hybridization (SSH, Diatchenko et al, 1996) using the PCR-Select cDNA subtraction kit (Clontech) according to the suppliers' instructions. cDNA

derived from the mutant line *mom1* is used as tester and cDNA derived from the parental line A as driver cDNA population. The subtracted library is cloned into vector pCR2.1 (Invitrogen). 500 individual bacterial cultures from this library are grown according to the manual of the PCR-Select differential screening kit (Clontech). To reduce the number of false positive clones, the library is primarily screened by inverted Northern blots as described by von Stein et al. (1997). Twelve among the 500 primarily selected cDNA clones show increased abundance upon hybridization with labeled *mom1* cDNA. Direct Northern blot analysis comparing total RNA of the wild type and the mutant line with these 12 cDNAs as probes reveal a striking genotype-dependent differential expression for two of them. Said clones are sequenced using conventional rhodamine or dRhodamine dye terminators from PE Applied Biosystems and a Perkin-Elmer GeneAmp PCR system 2400, 9600 or 9700 thermocycler. The sequence reactions are analyzed using an ABI PRISM 377 DNA sequencer. The cDNA clones are named TSI-A (903 bp, SEQ ID NO: 5) and TSI-B (614 bp, SEQ ID NO: 6). Both are abundant in the *mom1* RNA but are undetectable in Arabidopsis line A and wild type Arabidopsis. No consistent differential expression between mutant and wild type is observed for the remaining 10 cDNAs.

5' and 3' extension reactions are performed using Clontech's Marathon Kit according to the manufacturer's instructions. Sequence specific primers are

TA-F1: 5'-TGGTTCACCAGATAAGCTCAGTGCCCTC-3' (SEQ ID NO: 11) and

TA-F2: 5'-CTTCAGACTGGATAGGACTAGGTGGGCG-3' (SEQ ID NO: 12, nested primer),

for the 3'extension reaction and

TA-R1: 5'-CGCCCACCTAGTCCTATCCAGTCTGAAG-3' (SEQ ID NO: 13) and

TA-R2: 5'-CGCATCAAACAACTAACAACGAGGGCAC-3' (SEQ ID NO: 14, nested primer).

for the 5'extension.

PCR amplification products are cloned into vector pCR2.1 (Invitrogen). Individual bacterial cultures are grown and subjected to colony PCR as described in the manual of Clontech's PCR-Select Differential Screening Kit, with the primer combinations used to create the extension reactions (Marathon Adapter primer Ap1 (Clonetech) combined with TA-R2 or TA-F2 for the 5'- or 3' extension reactions, respectively). To screen for positive TSI-A extension clones, the PCR products are blotted and hybridized to TSI-A. All PCR reactions for cloning procedures are performed with a polymerase mix performing proofreading activity (Advantage cDNA PCR Kit, Clontech).

Since only two transcripts are detected in the polyadenylated RNA fraction of *mom1* plants, this RNA is used for 5' and 3' extensions reactions starting from the TSI-A sequence. Two clones each are analyzed at the nucleotide sequence level. The 5' extension yields inserts of 2512 bp (clone a, SEQ ID NO: 1) and 1997 bp (clone b, SEQ ID NO: 2), which after alignment are 97 % identical to each other. The clones from the 3' extension have a length of 1682 bp (clone c, SEQ ID NO: 3) and 1652 bp (clone d, SEQ ID NO: 4) and are 94 % identical. Interestingly, both 3' extension clones of TSI-A contain a region of 569 bp closely related to TSI-B (77 % identity). This explains the detection of similar RNA species on Northern blots with TSI-A and TSI-B as probes and their hybridization to the same YAC and BAC clones, and suggests that TSI-A and TSI-B are part of the same polyadenylated transcript species expressed in the *mom1* mutant. To confirm that the 5' extensions of TSI-A are indeed part of the TSI transcripts, a *mom1* Northern blot is probed with a cDNA fragment close to the 5' end of the extension (probe ORF corresponding to nucleotides 943-1334 of SEQ ID NO: 1). Interestingly, only the about 5000 nt long transcripts in the poly(A) fraction hybridize to this probe. Since this class of transcripts hybridizes to TSI-A and TSI-B, the 5000 nt transcripts are probably produced from templates containing a particular order of all three sequence elements.

Example 2: **Northern and Southern blot analysis and library screens**

Total RNA is either isolated as described by Goodall et al. (1990) or by the RNeasy Plant Mini Kit (Qiagen) according to the suppliers' instructions. For Northern blot analysis, the RNA is electrophoretically separated after denaturation by glyoxal in a 1.5 % agarose gel in phosphate buffer (pH 7) and blotted to nylon membranes (Hybond N, Amersham) using standard protocols. The Boehringer molecular weight marker I is used as a size standard. For Southern blot analysis, genomic DNA is isolated according to Dellaporta et al. (1983) and separated electrophoretically after endonucleolytic digestion. DNA fragments are transferred to nylon membranes (Hybond N, Amersham) according to standard procedures. Hybridization and washing of Northern and Southern blots and the filters with the YAC library is performed according to Church and Gilbert (1984). Probes are labeled with [α-^{32}P]-dATP by random prime DNA polymerization (Feinberg and Vogelstein, 1983) and exposed to X-ray sensitive film (Kodak X-OMAT AR).

Hybridization of Northern blots using total RNA prepared from 2-week-old seedlings with TSI-A of *mom1* visualizes four major transcripts with sizes of approximately 5000, 4700,

2500 and 1250 nucleotides. A TSI-B probe detects mainly two transcripts of 5000 nt and 2500 nt. Interestingly, the polyadenylated fraction of *mom1* RNA contains only the transcripts of 5000 nt and 2500 nt hybridizing to both cDNA probes (TSI-A and TSI-B). From the sizes of TSI transcripts detected it is obvious that both TSI clones represent only partial cDNAs. TSI expression is meiotically heritable and persists through 6 selfed generations of *mom1* with the same pattern of transcripts.

To examine TSI expression in other genotypes known to affect gene silencing, total RNA of several Arabidopsis mutant and transgenic strains known to be affected in gene silencing is probed with TSI-A. All the *som* mutants *som1* to *som8*, described by Mittelsten Scheid et al (1998) to be impaired in the maintenance of transcriptional silencing similar to *mom1*, show a high level of TSI-A expression. The mutation *ddm1*, originally identified to have decreased DNA methylation (Vongs et al. 1993), and later revealed to release transcriptional gene silencing from different loci (Mittelsten Scheid et al, 1998; Jeddeloh et al., 1998) also shows a high level of TSI-A expression. TSI-A expression is also expressed in a transgenic line described by Finnegan et al (1996) which shows decreased DNA methylation due to overexpression of DNA methyltransferase antisense mRNA as well as in a further Arabidopsis mutant affected in the DNA methyltransferase gene (said mutant is referred to as *ddm2* and has been provided by Eric Richards). Moreover, the silencing mutants *hog1* and *sil1*, but not *sil2* described by Furner et al (1998) express sequences hybridizing to TSI-A. Importantly, mutations affecting posttranscriptional gene silencing such as *sgs1* and *sgs2* described by Elmayan et al (1998) and egs1 described by Dehio and Schell (1994) do not express RNA which hybridizes to TSI-A.

Comparison of patterns of TSI-A expression in the different genotypes reveals genotype specific differences in the stochiometry of the different RNA species. *mom1* plants reveal a different expression pattern of TSI-A and TSI-B as compared to *som1* plants. These results indicate that a particular genetic deficiency in the transcriptional silencing system leads to a differential but specific activation of TSI templates. However, we observe variation in these activation patterns between different sources of plant material and different RNA preparations. Therefore, it is possible that patterns of TSI expression are more flexible and probably also controlled by still unknown factors acting in the mutant background, or by different stabilities among the transcript populations.

TSI-A and TSI-B are used as probes for Southern blots to determine the source of TSI transcripts and the organization of their template(s). The blots reveal that multiple copies of

TSI-A- and TSI-B-homologous sequences are present in the genome of Arabidopsis. Copy numbers are assessed by reconstruction experiments to approximately 130-300 copies of TSI-A.

To examine the degree of evolutionary conservation of the TSI arrangement, DNA of five Arabidopsis ecotypes (Zürich, Columbia, Landsberg erecta, Wassilevskija, C24) is compared by Southern blot analysis and hybridization to the TSI-A probe. Genomic DNA is digested with *Dra*I that has a single recognition site within TSI-A and *Ssp*I that does not cut within TSI-A. A significant conservation of the TSI-A pattern among different ecotypes is observed, with two main *Dra*I repeats of 4 kb and 1.3 kb and two major *Ssp*I fragments of 11 kb and 4 kb. Some minor differences specific for a particular ecotype indicate a limited genetic polymorphism within TSI-A. Probing the same membrane with TSI-B reveals complex banding patterns different in each ecotype which might indicate a lower degree of conservation for TSI-B, although the differences of the Southern blot patterns between TSI-A and TSI-B can also be explained if TSI-A is an internal part of a longer repeated element, and TSI-B is located proximal to a flank between repeated elements and variable single copy DNA regions.

After hybridizing TSI-A and TSI-B to the CIC YAC library covering 4 genome equivalents and 92% of the *Arabidopsis* genome sixty-two CIC clones out of 1152 turn out to hybridize with the TSI-A probe. Twenty-six of these contain also the pericentromeric 180-bp-repeat, 7 contain 5S RNA genes known to be located in the vicinity of a centromere, and 16 clones contain other markers that map close to centromeres. Only 4 of these clones map outside of centromeric regions. Similar mapping of TSI-B results in hybridization to all TSI-A positive CIC clones, with additional 7 clones hybridizing to TSI-B only. Thus both TSI repeats are concentrated in the pericentromeric regions of *Arabidopsis* chromosomes.

After hybridizing TSI-A and TSI-B to a cDNA library prepared with RNA isolated from *mom1* mutant plants, 22 hybridizing clones are further analyzed by sequence analysis. RNA is isolated from 2-week-old seedlings of the *mom1* mutant plant according to Goodall et al. (1990). The cDNA library is prepared using the Uni-ZAP XR library construction kit (Stratagene) according to the manufacturer's protocol. cDNA fragments larger than 500 bp are selected using the cDNA size fractionation columns from Gibco. 7 clones contain SEQ ID NO: 7 (TSI-A-15), 5 clones contain SEQ ID NO: 8 (TSI-A-2) part of which is identical to SEQ ID NO: 4.

Example 3: **Database searches**

Sequence analysis is performed using the GCG software (Wisconsin Package Version10.0, Genetics Computer Group (GCG), Madison, Wisconsin). For identity searches, GenEMBL, the Arabidopsis thaliana database (http://genome-www3.stanford.edu/), Kazusa Arabidopsis opening site (KAOS; http://zearth.kazusa.or.jp/arabi/) and Swissprot are used. Peptide sequences are analysed by GeneQuiz (http://columba.ebi.ac.uk:8765/gqsrv/submit) or Expasy (http://www.expasy.ch/).

Searches within the genomic sequence databases GenEMBL, Arabidopsis thaliana Database, and KAOS confirm the presence of multicopy sequences related to TSI-A and TSI-B, which are distributed over all five chromosomes of *Arabidopsis thaliana*. Importantly, very often single BAC clones contain sequences homologous to both cDNA clones. In some cases, TSI-A and TSI-B related sequences are found more than once on the same BAC clone, suggesting that TSI-A and TSI-B belong to a clustered repetitive element.

The significant sequence heterogeneity between the cDNA classes and duplicates of the 5' and 3' extensions of TSI-A indicate that they originate from different activated repeats. To facilitate the data base search for a possible genomic template of the 5000 nt transcript among the multiple related copies, the overlapping cDNA sequences are combined to form a continuous 4860 bp sequence of "virtual" cDNA (SEQ ID NO: 9) which is used to search the Arabidopsis genomic sequence databases. A particular BAC clone (TAMU BAC T6C20, accession number AC005898) has 91 % identity to the combined cDNA sequence. Further, the search uncovers a chromosomal DNA stretch (BAC F7N22, accession number AF058825) 99% identical to the abundant cDNA A-15 of Example 2 (SEQ ID NO: 7). The genomic sequence of the transcribed region 5' to the region defined by SEQ ID NO: 7 is given in SEQ ID NO: 27. It is identical to nucleotides 65081-68202 of BAC F7N22. Both sequences are located at the pericentromeric region of chromosome five. The TSI sequence defined by nucleotides 65080 to 70370 on BAC F7N22 is 54 % identical to the retrotransposon-like repeat named *Athila*. The identity of this sequence as a retrotransposon is deduced from *Arabidopsis* genome sequences around heterochromatic regions that are marked by the presence of 180 bp satellite repeats. The 10.5 kb sequence of *Athila* has several characteristics of a retroelement, like long terminal repeats (LTR), a polypurine track (PPT) and a primer binding site (PBS) for tRNA priming of the reverse transcriptase, but its open reading frames do not share homology with proteins known to be

involved in transposition.

The TSIs map to the 3' terminal part of *Athila*. TSI-A covers a part of the 3' non-coding region of the putative retrotransposon and TSI-B corresponds to the PPT and a part of the 3' LTR. The sequence of the 5' TSI-A extension encodes a possible open reading frame of 648 amino acids length (SEQ ID NO: 10) with 51 % identity to 604 amino acids of the ORF2 deduced for *Athila*. The sequence coding for this ORF is also present on the TAMU BAC T6C20, however, the ORFs encoded by the two cDNAs clone a (SEQ ID NO: 1) and clone b (SEQ ID NO: 2) and the BAC are interrupted by translational stop codons after 398 amino acids (clone a), 83 amino acids (clone b) and 46/465/496/499/549 amino acids respectively (BAC T6C20). The ORF2 sequence present on BAC F7N22 is highly degenerated by five deletions of 2-31 bp and five insertions of 3-10 bp. This further supports the assumption that this sequence is derived from a putative but degenerated retrotransposon. Data base searches for proteins similar to the potential product of the 648 amino acids ORF do not yield significantly similar polypeptides neither to proteins usually encoded by retroelements nor to any other known polypeptides.

Example 4: **RNAse protection assays**

RNase protection assays are performed according to Goodall et al. (1990) with minor modifications. To assay the direction of TSI transcription, the pCR2.1 based plasmid containing the TSI-A insert is cut by EcoRI creating a fragment of 781 bp which is ligated into the vector pGEM-7Zf(+) (Promega). To map the 5' transcription start, the probe is generated by amplifying the BAC F7N22 region between positions 64929 and 65567 and inserting the product into the pGEM-7Zf(+) vector (Promega). Labeled probes are synthesized by *in vitro* transcription of the linearized plasmid in the presence of [α-^{32}P]-UTP using T7 polymerase (Promega) or Sp6 polymerase (Boehringer) and purified by electrophoresis (Goodall et al., 1990). Single stranded RNA is cleaved by either 4 µg RNase A and 0.6 U RNase T_1 (RNase A/T assay) or by 20 U RNase T_1 (RNase T assay). Protected fragments are separated on a denaturing 6 % polyacrylamide gel. The dried gel is exposed to a PhosphorImager screen (Molecular Dynamics) and to X-ray sensitive film.

To determine the polarity of TSI transcription, RNase T and RNase A/T protection assays are performed with TSI-A probes of opposite polarity. TSI-A sequences are used as probes since TSI-A is present in all transcripts detected on Northern blots. There is no evidence for protection of the probe corresponding to the sense strand. This suggests the lack of any

TSI antisense RNA and a unidirectional transcription of the TSI templates. Interestingly, RNase digestions with a TSI-A antisense probe creates a complex pattern of TSI-A protected bands. This suggests that many different but related RNAs hybridize to the probe. Since a fragment of 781 nt as expected for the protection of the entire TSI-A probe is visible, it can be concluded that TSI-A is part of an activated transcript throughout and not an artifact generated by template switch during the SSH procedure. Furthermore, some of the protected TSI-A fragments are clearly more abundant than others, suggesting either a structural conservation of particular regions of TSI-A within related RNAs, or alternatively a higher abundance of certain transcript subspecies.

The sequence information of BAC F7N22 is used to determine the position of the transcription start for the longest TSI transcript. An antisense RNA probe for RNase A/T protection is produced spanning the 638 nucleotides between positions 64929 and 65567. The probe is hybridized with total RNA from *ddm1*, *som7* and *mom1*. In all RNA preparations, a fragment of approximately 480 nt (±10 nt) is protected and allowes positioning of the TSI transcription start on BAC F7N22 to 65087 (± 10 nt) in different mutants.

Example 5: **Reverse Transcription PCR (RT PCR)**

Reverse transcription is performed with 1 µg total RNA from *mom1* in the presence of 1 mM dNTPs, 4-20 U RNasin (Promega), 1x AM RTase buffer (Boehringer) and 25 U AM reverse transcriptase (Boehringer) at 37°C for 1 hour, followed by heat inactivation of the reaction mixture. As template for PCR 50 ng reverse transcribed RNA primed by gene specific antisense primers (BA-R1, BA-R2, AT-R1, and TA-R1, see below) or 100 ng genomic DNA or 100 ng cDNA are used. PCR is started with 3 min denaturation at 94°C, followed by 30 amplification cycles (denaturation at 94°C/30 sec, annealing at 62°C/30 sec, and elongation at 72°C/30 sec) in the presence of 0.2 mM dNTPs, 0.4 µM forward and reverse primers, 1x Taq DNA polymerase buffer (Boehringer) and 0.25 U Taq DNA polymerase (Boehringer). The nucleotide sequences of the primers used for RT-PCR are:

AT-F1: 5'-CGATAACATCGACCGTATTGCTCGCC-3' (SEQ ID NO: 15)

AT-R1: 5'-AACTAGCTCCCATCCGTCTTCGACATCC-3' (SEQ ID NO: 16)

AT-F2: 5'-TGCATCACACCGGATTGGATTGAC-3' (SEQ ID NO: 17)

AT-R2: 5'-TGTTCCCCTGAACCATAGCAATGAGACC-3' (SEQ ID NO: 18)
BA-F1: 5'-CAAACAGACAGAGTGTGGCCCACCACC-3' (SEQ ID NO: 19)
BA-R1: 5'-AAGAGAGGGAGAAGGCAGTGGCGTGAG-3' (SEQ ID NO: 20)
BA-F2: 5'-TGCAAACCCACAGGACCAAGTCTACCC-3' (SEQ ID NO: 21)
BA-R2: 5'-ACAGATGGTGATAGCGTGAGCGGTGGC-3' (SEQ ID NO: 22)
F7-F1: 5'-TCAACCTTTTGCCCCAACAACCACTC-3' (SEQ ID NO: 23)
F7-R1: 5'-TCTCCATCCACGCTTTCCTGAATGTCC-3' (SEQ ID NO: 24)
GS-F1: 5'-GGAGAAGGAAGCTGAAAATCATATTGTGG-3' (SEQ ID NO: 25)
GS-R1: 5'-ATGATGATCCTAAGTCTACCCTTTTGCAC-3' (SEQ ID NO: 26)

As a positive control for the PCR reactions, the TA-F1 and TA-R1 primers are used. The reverse transcriptase region of the pol gene of Arabidopsis Ty1/copia-like retrotransposon family is amplified as described by Konieczny (1991) with immaterial modifications.

Since the nucleotide sequence of the TSI transcripts TSI-A and TSI-B is related to the nucleotide sequence of the 3' half of retrotransposon-like element *Athila* including the second ORF and the 3' LTR, we examined by RT-PCR whether transcription of the 5' part of *Athila* including the first ORF is activated in *mom1* plants. Five primer pairs are chosen (BA-F1 - BA-R1; BA-F2 - BA-R1; AT-F1 - AT-R1; AT-F2 - AT-R2, F7-F1 - F7R1) according to the sequence information about *Athila* and the related parts of BAC T6C20 and BAC F7N22. All primer combinations amplified the expected products from genomic template DNA, but no PCR product could be obtained from *mom1* RNA, regardless, whether cDNA synthesis was started from an *Athila*- or BAC-specific reverse primer or from polyT-primed cDNA (data not shown). Activation of TSI therefore is limited to sequences related to the 3' part of *Athila*.

The two classes of isolated cDNAs share only approximately 50% identity with *Athila*. To directly address the question of whether *Athila* is expressed, RT-PCR experiments are performed with *Athila*-specific primers (GS-F1, GS-R1) in the TSI homologous region. However, the corresponding fragment cannot be amplified from RNA of *mom1* seedlings, suggesting that only a subset of *Athila*-like sequences but not the *Athila* element itself is reactivated in the mutant background.

To investigate whether other retroelements are transcriptionally activated in *mom1*, degenerated primers in a conserved region of the reverse transcriptase gene used to clone

and describe the Ta superfamily of Arabidopsis retrotransposons (Konieczny et al., 1991) are used to investigate, if members of the Ta family are transcribed in the mutant background. Although the expected 268 bp fragment can be amplified from genomic DNA, no amplification is achieved in RT-PCR with *mom1* RNA as template. This indicates, that, in spite of the TSI homology to retrotransposons, these elements are not generally activated in the *mom1* mutant.

Example 6: **TSI expression after application of stress (salinity, UV-C, pathogen)**

Induction of TSI upon UV-C is tested on Northern blots with RNA samples from 1 week-old-seedlings subjected to UV-C treatment of 1 kJ/m^2 or 5 kJ/m^2 which are collected at several time points within 1 hour (Revenkova et al., 1999). The effect of osmotic stress is tested on Northern blots with RNA from one-week-old seedlings that are transferred for 24 hours to medium with NaCl concentrations of 0, 0.04, 0.08 and 0.12 M (Albinsky et al., 1998). To test TSI expression upon pathogen stress, RNA of 3-week-old seedlings either mock treated or infected with *Peronospora* is analysed by Northern blot analysis. To verify the appropriate pathogen response, induction of PR1 expression is monitored by reprobing the membrane with a PR1 probe.

In young seedlings (2 weeks old) and in different tissues of mature wild type plants (roots, shoots, leaves, flowers, siliques), TSI expression cannot be detected. The application of various stress treatments namely elevated salinity, UV-C, or pathogen infection, does not activate TSI in wild type plants. TSI expression is also not detected in freshly initiated callus cultures, and transcriptional suppression of TSI is stable even after several *in vitro* passages of the callus culture. However, the only exception so far are cells derived from wild type *Arabidopsis* (literature) growing for a long time in suspension culture. These cells express TSI-A, indicating release of TSI silencing under these conditions.

REFERENCES

- Albinsky et al, *Plant J.* **17**: 73-82, 1999
- Amedeo et al, *Nature* **405**: 203-206, 2000
- Church and Gilbert, *Proc Natl Acad Sci USA* **81**: 1991-1995, 1984
- Dehio and Schell, *Proc Natl Acad Sci USA* **91**: 5538-5542, 1994
- Dellaporta et al, *Plant Mol Biol Rep.* **1**: 19-21, 1983
- Diatchenko et al, *Proc Natl Acad Sci USA* **93**: 6025-6030, 1996
- Dilkes and Feldmann, *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 339-351, 1998
- Elmayan et al, *Plant Cell* **10**: 1747- 1757, 1998
- Feinberg et al, *Anal Biochem.* **132**: 6-13, 1983
- Feldmann et al, *in* Arabidopsis (Ed. Sommerville and Meyerowitz), CSHL Press, New York, pp 137-172, 1994
- Finnegan et al, *Proc Natl Acad Sci USA* **93**: 8449-8454, 1996
- Furner et al, *Genetics* **149**: 651-662, 1998
- Goodall et al, *Methods Enzymol.* **181**: 148-161, 1990
- Jeddeloh et al, *Genes Dev.* **12**: 1714-1725, 1998
- Konieczny et al, *Genetics* **127**: 801-809, 1991
- Long and Coupland *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 315-328, 1998
- Mittelsten Scheid et al, *Proc Natl Acad Sci USA.* **95**: 632-637, 1998
- Pellisier et al, Plant Mol Biol. 29: 441-452, 1995
- Pereira and Aarts, *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 329-338, 1998
- Revenkova et al, *EMBO J.* **18**: 490-499, 1999
- von Stein et al, *Nucleic Acids Res.* **25**: 2598-2602, 1997
- Vongs et al, *Science.* **260**: 1926-1928, 1993